DOCUMENT RESUME

ED 036 247 HE 001 232

AUTHOR LUNNEBORG, CLIFFORD E.
TITLE DIRECT VALIDATION OF DIFFERENTIAL PREDICTION.
INSTITUTION WASHINGTON UNIV., SEATTLE. BUREAU OF TESTING.
PUB DATE SEP 69
NOTE 15P.; RESEARCH REPORTED AT A PSYCHOMETRIC CONFERENCE HONORING PAUL HORST, SEATTLE, WASHINGTON, JUNE 22, 1969

EDRS PRICE EDRS PRICE MF-$0.25 HC-$0.85
DESCRIPTORS *ACHIEVEMENT TESTS, COMPARATIVE TESTING, *HIGHER EDUCATION, PREDICTIVE ABILITY (TESTING), *PREDICTIVE VALIDITY, *STUDENTS, *TESTING

ABSTRACT

USING ACADEMIC ACHIEVEMENT DATA FOR 655 UNIVERSITY STUDENTS, DIRECT VALIDATION OF DIFFERENTIAL PREDICTIONS BASED ON A BATTERY OF APTITUDE/ACHIEVEMENT MEASURES SELECTED FOR THEIR DIFFERENTIAL PREDICTION EFFICIENCY WAS ATTEMPTED. IN THE CROSS-VALIDATION OF THE PREDICTION OF ACTUAL DIFFERENCES AMONG FIVE ACADEMIC AREA GPA'S, THIS SET OF DIFFERENTIAL PREDICTORS HAD ONLY A SLIGHT ADVANTAGE OVER A SET SELECTED TO MAXIMIZE ABSOLUTE PREDICTION OF THE FIVE AREA GPA'S. THREE WAYS IN WHICH THE EFFECTIVENESS OF DIFFERENTIAL PREDICTION CAN BE IMPROVED ARE PRESENTED. (AUTHOR)

N-X

Bureau of Testing

University of Washington

September 1969

ED036247

Direct Validation of Differential

Prediction[1]

Clifford E. Lunneborg

Using academic achievement data for 655 university students direct validation of differential predictions based on a battery of aptitude/achievement measures selected for their differential prediction efficiency was attempted. In the cross-validation of the prediction of actual differences among five academic area GPA's this set of differential predictors had only a slight advantage over a set selected to maximize absolute prediction of the five area GPA's. Three ways in which the effectiveness of differential prediction can be improved are presented.

HE 001 232

Bureau of Testing Project: 126

[1]Research reported at a Psychometric Conference Honoring Paul Horst, Seattle, June 22, 1969.

Bureau of Testing

University of Washington

September 1969

Direct Validation of Differential Prediction

Clifford E. Lunneborg

In a pair of monographs authored in the middle fifties, Horst (1954, 1955) drew an important distinction between two possible goals associated with the selection of predictor variables where success in a number of criterion activities is to be predicted. On the one hand predictors may be selected which together will have the greatest predictive efficiency over all of the criterion activities. As usually interpreted this means selecting a set of predictors, of a given size, which will yield the highest average multiple correlation over the several criteria, and Horst (1955) has referred to such a selection strategy as maximizing multiple <u>absolute</u> prediction. By contrast, a set of predictors, again of some given size, may be selected so as to maximize the efficiency with which the battery on the average predicts differences among the criterion activities. This second strategy maximizes multiple <u>differential</u> prediction (Horst, 1954).

The two strategies differ not only by selecting different predictors but in the appropriateness of these selections for two contrasting decision situations. It is reasonable to expect the absolute prediction strategy to select as predictors variables highly related to whatever is common among the criterion activities, and the differential strategy to search for measures correlated with differences between criterion activities.

To adopt the distinction suggested by Cronbach and Gleser (1965) absolute prediction would seem the better model for institutional decision making and differential prediction more appropriate for individual decisions. The institution's requirement most commonly is to choose among individuals. Individuals are selected for assignment to a criterion activity if they are expected to do better in that activity than other available individuals. To accomplish this it is desirable to have the most accurate estimate for each individual of his success in each criterion activity. For individual decisions, on the other hand, the choice is among criterion activities. It is more important for the individual to have accurate information about whether he can expect to do better in one activity compared with another activity than to know with exactness how well he might do in any one of the activities.

A characteristic common to both decision situations is the incompleteness of data relating to criterion activities. Where criterion activities are numerous it is frequently impossible to obtain performance data for all activities for any individual. For example, in the university setting where course work is available in fifty distinct areas, the typical student in four years only samples eleven to fifteen areas. Because choice of criterion activities whether by the individual or the institution is not independent of the individual's predictor data, the incompleteness of criterion data clearly poses theoretical problems in the estimation of regression parameters. In practice, however, these problems are often ignored (Johnson, 1959). After all, the proof of the goodness of prediction is easily accomplished through cross-validation, at least as far as absolute prediction is concerned. One straightforwardly calculates correlations between predicted and actual criterion performances for some subsequent sample and compares these correlations with the multiple correlations obtained when variables were selected.

Unfortunately, this has also been the typical means of cross-validating differential predictions. Although Horst's (1954) differential predictor selection technique is keyed to maximizing the prediction of differences between all pairs of criteria, his solution does not depend upon the availability of such criterion differences or even of correlations between criteria. Rather, predictor selection requires only "adequate" estimates of the intercorrelations of all the potential predictors and of the correlation of each of these measures with each criterion. Granting the adequacy of these estimates, differential predictor selection can thus proceed with quite incomplete criterion data. However, using these selected differential predictors and validating their use raises issues not faced with absolute prediction. The problem is that once variables are selected as differential predictors they tend to be used and "validated" as if they were absolute predictors. To predict the difference between two criteria the appropriate weighting for a predictor is simply the difference between the weights that predictor would have if each of the two criteria were to be predicted separately. Thus, the best prediction in the least squares sense of the difference between achievement in sociology and biology, using a set of selected predictors, is obtained by weighting each of these predictors to obtain the best least squares predictions separately of sociology and biology and then taking the difference between these weightings. This simple relationship together with the greater ease of interpreting predictions for k criteria rather than [(k/2) x (k-1)] differences between criteria, leads advocates of differential prediction to make the differential aspect of predictions more implicit than explicit, e.g., in the Washington Pre-College (WPC) Testing Program (WPC, 1969b). In this program success in each of many academic areas is predicted on the basis of variables selected to differentiate among these areas and it is left to the user of these predictions, the high school student and his counselor, to make the comparisons among predictions and thus extract the differential information. The question of how best to present differential predictions certainly has not been adequately studied, but that question goes beyond the concern of the present paper.

However, once differential predictors have been used not to provide direct predictions of differences but instead predictions of the individual criteria to be differentiated by the user, it is easy to see why validation of such predictions has rested on techniques developed for the absolute model. To cite from the WPC Testing Program again, weights are derived for the selected differential predictors to best estimate each of the "differentiable" criteria. These weights are then applied to predictor data for new samples and these prediction equations "validated" by correlating the resulting predicted achievements with obtained achievements (WPC, 1969a). Such a strategy validates the accuracy of prediction of the individual criteria but it does not answer the question of how well differences among criteria have been predicted. So far no attempt has been made to validate differential prediction in this latter sense.

Method

The present study demonstrates the direct validation of differential prediction for a limited sample of students entering the University of Washington (UW) as freshmen after having completed the WPC battery. As a part of a larger study 655 students were identified who entered UW fall 1966 and took through spring 1968 a minimum of two courses in each of five areas of study: freshman English, college algebra, modern foreign languages, natural sciences, and social sciences. Requiring that each student had an earned grade point average (GPA) in each of the five areas meant that differences in performance could actually be computed for all ten pairings of achievement areas. This group of Ss comprised 26% of the tested freshman class, the remainder of which did not have the above pattern of course work.

The sample was split into halves by assigning alternate students from an ordering based on the serially assigned student admissions number to one of two groups. A complete intercorrelation matrix for the twenty WPC predictor variables identified in Table 1, the five college GPA's and ten GPA differences was calculated separately for the two groups. Group I (N = 328, 39% female) correlations were utilized for predictor selection and Group II (N = 327, 41% female) data for cross-validation.

Two predictor selections were conducted on Group I correlational data. Six predictors were selected from the twenty to maximize absolute prediction for the five area GPA's. Then, in the second selection, six predictors were selected to maximize differential prediction of these same five academic areas. In both instances Horst's sequential selection strategies were employed. For the selection of absolute predictors that measure was first selected which had the highest average correlation with the five criteria (Horst, 1955). The second selected predictor was the one which, combined with the first, produced the highest average multiple correlation with the five criteria. Additional predictors were then selected one at a time in accordance with this rule. The selection of six differential predictors proceeded in the same fashion except that the function to be maximized at each stage was the average multiple correlation across all possible differences between the five GPA's. According to Horst (1954) this was equivalent to maximizing the difference between the average variance of the five predicted GPA's and the average of their covariances.

Following each of these selections two predictor weightings based on Group I data were calculated. Weights were first derived for the members of each predictor set to maximize the efficiency with which each of the five college GPA's could be estimated. Secondly, the two sets of selected predictors were weighted to predict each of the ten differences between these GPA's by subtracting the weight appropriate to one of the area GPA's from the weight for the other in the pair. For example, weights were derived to predict the difference, "English GPA minus Algebra GPA" by subtracting from the weights for predicting English GPA those for predicting Algebra GPA. Both sets of weights were then applied to the predictor data for Groups I and II and correlations calculated between predicted and actual GPA's and GPA differences.

Table 1

Order of Selection of Predictors

Absolute Predictors			Differential Predictors		
Predictor	Δ	λ	Predictor	Δ	ϕ
English Usage	.688	.688	Vocabulary	.070	.070
HS Nat Sci GPA	.195	.883	Spelling	.037	.107
Reading Comp	.153	1.036	HS Math GPA	.034	.141
Math Achieve	.129	1.164	HS English GPA	.038	.179
Sex	.071	1.236	HS Nat Sci GPA	.026	.205
HS For Lang GPA	.050	1.286	English Usage	.022	.227

Note.--Variables unselected included HS Social Studies GPA, HS Electives GPA, Age, Reading Speed, Mechanical Reasoning, Spatial Ability, Applied Mathematics, Quantitative Judgement, Data Sufficiency, and Functional Relationships.

Results

Table 1 indicates the predictors selected for Group I. Predictor variables are listed in the order in which they were selected. For the absolute predictor selections λ is Horst's (1955) index of absolute prediction efficiency, the sum of the multiple correlations between predictors selected and the five area GPA's. The Δ elements are the increments to λ with each predictor selection. Similarly, for the differential predictor selections ϕ is Horst's index of differential prediction efficiency (1954), proportional to the difference between the average variance and the average covariance of the area GPA's predicted at each stage. Increments to this index are again tabled as Δ. The differential index, ϕ, does not have the same interpretation as the index for absolute prediction. ϕ can never exceed the difference between unity and the average intercorrelation of the area GPA's. In this study these intercorrelations were known and, for Group I, the resulting maximum was .457. Only two predictors, English Usage and HS Nat Sci GPA, were common to the two sets of selections.

Table 2 reports the squared multiple correlations, R^2, of the two selected sets of predictors with each of the five GPA areas for Group I data. Also reported are squares of correlations between predicted and actual GPA's for Group II. These latter are tabled as $r^2_{Y\hat{Y}}$. The battery of absolute predictors had the higher index of absolute prediction efficiency, 1.29 as opposed to 1.08, but striking was the shrinkage of predictability associated with the last three GPA areas. On cross-validation the two sets of predictors provided essentially the same level of absolute prediction.

Table 2

Multiple Correlations and Cross Validations

for Five College Course Area GPA's

Area GPA's	Absolute Predictors R^2	Absolute Predictors $r^2_{Y\hat{Y}}$	Differential Predictors R^2	Differential Predictors $r^2_{Y\hat{Y}}$
English	.2997	.2117	.3146	.2534
Algebra	.1754	.0950	.1193	.0801
Foreign Languages	.2716	.1719	.2481	.1658
Natural Sciences	.2602	.1019	.2012	.1073
Social Sciences	.2793	.1018	.1934	.1004
Sum	1.2862	.6823	1.0765	.7070
Average percent variance	25.7	13.6	21.5	14.1

Table 3 displays .. results for the predictions of differences between the college area GPA's. Differential prediction efficiency, as indexed here by the sum of the squared multiple correlations, was only slightly better for the differential battery, .87 versus .85. This advantage was, however, maintained in cross-validation. Also displayed in Table 3 are the correlations between the two GPA's contributing to each difference. The size of these college area GPA correlations was not strongly related to the predictability of the corresponding differences but, interestingly enough, the least correlated areas, English and Algebra, did provide the most highly predictable GPA difference.

Examining the cross-validation results, the absolute predictors accounted for about 13% of criterion variability, averaging across the five GPA's (Table 2), while the differential predictors accounted for only one-half that amount (6.5%) of the variance in the ten GPA differences (Table 3). Before discounting this differential prediction effort, it should be noted that reliability of GPA differences was undoubtedly a limiting factor. Given the correlations between the area GPA's and assuming that the reliabilities of these course area GPA's were as high as .70 or .80, then the GPA differences would only have had reliabilities of approximately .50. There was, then, considerably less reliable variance to be accounted for in GPA differences than in the simple GPA's.

Discussion

What are the implications for the development of differential prediction? The goal, identifying variables which can be used to improve individual decision making, is obviously an important one. Although the

Table 3

Multiple Correlations and Cross Validations

for Ten College GPA Differences

Difference GPA's	Absolute Predictors R^2	Absolute Predictors $r^2_{Y\hat{Y}}$	Differential Predictors R^2	Differential Predictors $r^2_{Y\hat{Y}}$	GPA r (Group I)
English - Algebra	.1229	.0997	.1348	.1090	.31
English - For Lang	.0894	.0865	.1277	.0937	.45
English - Nat Science	.1049	.0616	.1084	.0748	.40
English - Soc Science	.0682	.0123	.0342	.0019	.42
Algebra - For Lang	.0847	.0808	.0900	.0851	.38
Algebra - Nat Science	.0654	.0305	.0511	.0273	.45
Algebra - Soc Science	.0895	.0490	.0832	.0850	.34
For Lang - Nat Science	.0917	.0936	.1030	.0848	.52
For Lang - Soc Science	.0872	.0593	.0753	.0420	.46
Nat Sci - Soc Science	.0471	.0269	.0586	.0465	.57
Sum	.8510	.6002	.8663	.6501	
Average percent variance	8.5	6.0	8.7	6.5	

differential predictor selection strategy led in this study to slightly better prediction of GPA differences, the overall accuracy was very limited. Can this accuracy be improved? The answer is yes if three separate aspects of the differential prediction situation can be properly investigated. These are the composition of the predictor pool, the matching of the potential predictors with relevant differences among the criteria, and the selection of appropriate criteria.

To date, differential prediction efforts have largely depended upon the use of predictors which have established validity in the absolute prediction sense. For example, all of the WPC predictors are appreciably correlated with success in most traditional academic course areas. There is reason to believe that good differential predictors may not be like these. A recent prediction study with this same battery suggested that augmenting these aptitude and achievement measures with biographical and interest variables considerably improved differential prediction (Lunneborg, 1968). Although these new variables were selected as differential predictors, they were not picked when the goal was maximizing absolute prediction. Increased variety in the predictor pool, then, is one way in which differential prediction can be improved.

The variety of predictors, however, is not sufficient to establish their usefulness. Clearly, predictors must be related to what is different among criteria. In this study, that the most predictable difference was between English and Algebra GPA's may be understandable. The WPC battery measures, largely, verbal and quantitative skills. These skills should be useful in distinguishing subsequent English and Algebra achievements. If the difference between English and Social Sciences GPA's is to be predicted, variables must be found which are similarly differentially related to success in these areas.

Finally, and perhaps most importantly, differential prediction depends upon criteria being differentiable. If two criteria, within the constraints of their respective reliabilities, measure the same thing, it is useless to attempt a search for measures which will predict the difference between them. There is evidence in this study, for example, that the correlation between success in Natural Science and success in Social Science, .57, may rival in magnitude the reliabilities of these success measures. If this is true, then the difference between Natural and Social Science GPA's is a measure with essentially no reliable component to its variance and is necessarily unpredictable. It is not unlikely that there are a number of liberal arts achievement areas which cannot be differentiated. The prospect, in the WPC program, of contrasting selected liberal arts achievement areas with vocational-technical training areas will undoubtedly provide criterion differences which are sufficiently reliable as to make the search for varied and rationally related predictors worthwhile.

References

Cronbach, L. J., & Gleser, G. C. Psychological tests and personnel decisions. Urbana: University of Illinois Press, 1965.

Horst, P. A technique for the development of a differential prediction battery. Psychological Monographs, 1954, 68 (9, Whole No. 380).

Horst, P. A technique for the development of a multiple absolute prediction battery. Psychological Monographs, 1955, 69 (5, Whole No. 390).

Johnson, R. M. Differential prediction with incomplete criterion data. Unpublished doctoral dissertation, University of Washington, 1959. Mic. 60-859.

Lunneborg, C. E. Biographic variables in differential vs. absolute prediction. Journal of Educational Measurement, 1968, 5, 207-210.

Washington Pre-College Testing Program. Counselor's manual, 1969-70. Seattle: WPC, University of Washington, 1969. (a)

Washington Pre-College Testing Program. Student instruction booklet, 1969-70. Seattle: WPC, University of Washington, 1969. (b)